

The counter-propagating Rossby-wave perspective on baroclinic instability. I: Mathematical basis

Article

Published Version

Heifetz, E., Bishop, C. H., Hoskins, B. J. and Methven, J. (2004) The counter-propagating Rossby-wave perspective on baroclinic instability. I: Mathematical basis. Quarterly Journal of the Royal Meteorological Society, 130 (596). pp. 211-231. ISSN 1477-870X doi: <https://doi.org/10.1002/qj.200413059610> (Part A) Available at <https://centaur.reading.ac.uk/88/>

It is advisable to refer to the publisher's version if you intend to cite from the work. See [Guidance on citing](#).

To link to this article DOI: <http://dx.doi.org/10.1002/qj.200413059610>

Publisher: Royal Meteorological Society

All outputs in CentAUR are protected by Intellectual Property Rights law, including copyright law. Copyright and IPR is retained by the creators or other copyright holders. Terms and conditions for use of this material are defined in the [End User Agreement](#).

www.reading.ac.uk/centaur

CentAUR

Central Archive at the University of Reading

Reading's research outputs online

The counter-propagating Rossby-wave perspective on baroclinic instability. I: Mathematical basis

By E. HEIFETZ¹, C. H. BISHOP², B. J. HOSKINS³ and J. METHVEN^{3*}

¹*Tel-Aviv University, Israel*

²*Naval Research Laboratories/UCAR, Monterey, USA*

³*University of Reading, UK*

(Received 10 September 2002; revised 12 June 2003)

SUMMARY

It is shown that Bretherton's view of baroclinic instability as the interaction of two counter-propagating Rossby waves (CRWs) can be extended to a general zonal flow and to a general dynamical system based on material conservation of potential vorticity (PV). The two CRWs have zero tilt with both altitude and latitude and are constructed from a pair of growing and decaying normal modes. One CRW has generally large amplitude in regions of positive meridional PV gradient and propagates westwards relative to the flow in such regions. Conversely, the other CRW has large amplitude in regions of negative PV gradient and propagates eastward relative to the zonal flow there. Two methods of construction are described. In the first, more heuristic, method a 'home-base' is chosen for each CRW and the other CRW is defined to have zero PV there. Consideration of the PV equation at the two home-bases gives 'CRW equations' quantifying the evolution of the amplitudes and phases of both CRWs. They involve only three coefficients describing the mutual interaction of the waves and their self-propagation speeds. These coefficients relate to PV anomalies formed by meridional fluid displacements and the wind induced by these anomalies at the home-bases. In the second method, the CRWs are defined by orthogonality constraints with respect to wave activity and energy growth, avoiding the subjective choice of home-bases. Using these constraints, the same form of CRW equations are obtained from global integrals of the PV equation, but the three coefficients are global integrals that are not so readily described by 'PV-thinking' arguments. Each CRW could not continue to exist alone, but together they can describe the time development of any flow whose initial conditions can be described by the pair of growing and decaying normal modes, including the possibility of a super-modal growth rate for a short period. A phase-locking configuration (and normal-mode growth) is possible only if the PV gradient takes opposite signs and the mean zonal wind and the PV gradient are positively correlated in the two distinct regions where the wave activity of each CRW is concentrated. These are easily interpreted local versions of the integral conditions for instability given by Charney and Stern and by Fjørtoft.

KEYWORDS: Orthogonality PV gradient Wave interaction Zonal propagation

1. INTRODUCTION

Bretherton (1966b) and Hoskins *et al.* (1985), suggested a potential-vorticity (PV) based description of baroclinic/barotropic instability in terms of the zonal propagation and mutual interaction of two Rossby waves on a zonal shear flow (see Fig. 1). There is likely to be significant continuing interaction if the phase speeds of the two waves, in the absence of interaction, are similar. Since propagation relative to the zonal flow (hereafter shortened to 'propagation') has opposite sign to the basic-state meridional PV gradient (hereafter, 'PV gradient') it is anticipated that the stronger eastward zonal flow must be associated with a positive PV gradient to slow the wave in that region. Similarly, the weaker (or negative) zonal flow might be expected to be associated with a negative PV gradient to make the wave there propagate eastwards, thus achieving a similar phase velocity (advection plus propagation) to the first wave. This qualitative argument of 'counter-propagating Rossby waves' (CRWs), encapsulates the two signed PV-gradient baroclinic instability condition of Charney and Stern (1962), and the correlation of the mean zonal wind U and the mean PV-gradient condition of Fjørtoft (1951).

If the relative phases of the two waves are specified, then the sense of the interaction between the CRWs can be ascertained following the arguments by Hoskins *et al.* (1985).

* Corresponding author: Department of Meteorology, University of Reading, PO Box 243, Earley Gate, Reading, Berkshire RG6 6BB, UK. e-mail: J.Methven@reading.ac.uk

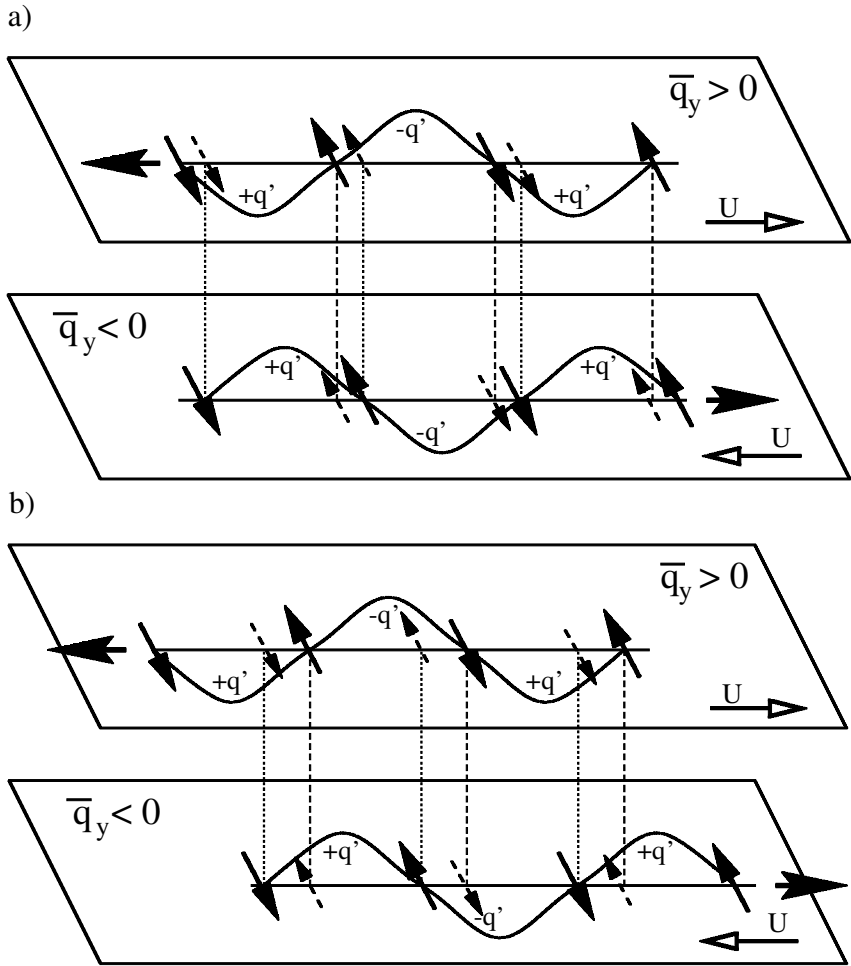


Figure 1. Schematic illustration of the upper and lower counter-propagating Rossby waves (CRWs) on a baroclinic basic state where \overline{q}_y is the mean potential-vorticity (PV) gradient and the mean wind, $U(z)$, relative to its average is indicated by the blank arrows. The phase of each CRW is indicated by its PV anomalies, q' , and each wavy line indicates the corresponding meridional displacements of a PV contour. The solid zonal arrows indicate the propagation direction of each CRW (counter to the relative zonal flow). The other solid arrows indicate the meridional wind associated with each CRW at its own level and the dashed arrows indicate the wind induced at the level of the other CRW. In (a) the PV anomalies of the upper CRW are shifted to the west of the lower CRW's PV anomalies by a phase between 0 and $\pi/2$. Consequently, the CRWs 'help' each other to counter-propagate and to grow. In (b) the phase difference between the upper and lower CRW is between $\pi/2$ and π . Consequently the CRWs still grow mutually but 'hinder' each other's counter-propagation.

For discussion purposes we take the pure baroclinic case $U(z)$ where the zonal flow increases with height and is shown relative to a mid-level value in Figs. 1(a) and (b). An upper CRW exists on a positive PV gradient and a lower CRW on a negative PV gradient. In Fig. 1(a) the upper PV wave is shifted to the west of the lower PV wave by less than $\pi/2$. The strongest poleward flow, and hence negative PV tendency, induced by the lower wave at the level of the upper wave is between $\pi/2$ and 0 to the west of the upper-wave negative PV extremum. It therefore amplifies the upper wave and enhances the propagation westwards. Similarly, the poleward flow (and hence the positive PV tendency induced by the upper wave at the level of the lower wave) is between 0 and

$\pi/2$ to the east of the lower-wave positive PV. This leads to growth and more rapid propagation eastwards. The interaction of the two CRWs leads to growth of each and makes each propagate more rapidly against the relative zonal flow. As each wave is assisting the other in its counter-propagation, this is referred to as ‘helping’.

For a westward shift with height between $\pi/2$ and π (Fig. 1(b)), the negative PV tendency induced by the lower wave at the level of the upper wave is between 0 and $\pi/2$ to the east of the upper wave negative PV. There is again growth, but this time the upper CRW’s westward propagation is reduced by the interaction. Similarly, it can be argued that the lower CRW grows and propagates less rapidly to the east as a consequence of its interaction with the upper wave. Thus, the interaction of the two CRWs again leads to growth of each but ‘hinders’ their counter-propagation (against the relative zonal flow).

Since the interaction acts to increase the westward phase shift when it is between 0 and $\pi/2$ and decrease it when it is between $\pi/2$ and π , then if the interaction is strong enough it can be expected to lead to the phase remaining in the range of $0 \rightarrow \pi$ and to continued growth. The equilibrium position for the phase would correspond to the growing normal mode. For eastward phase shifts, similar arguments apply and there is decay. However, in this case the hindering/helping mechanism tends to shift the phase from the decaying range of $-\pi \rightarrow 0$ toward the growing range of $0 \rightarrow \pi$. This indicates the non-modal evolution mechanism in the initial value problem which leads towards convergence to the phase-locking configuration of the growing normal mode.

This appealing picture of baroclinic instability was first proposed by Bretherton (1966b) in the context of the two-layer model. In this case, it is a direct physical description of the two layers. Following Bretherton’s argument, boundary potential temperature can be considered as a δ -function in PV: a positive potential-temperature anomaly on the lower boundary acting as positive PV, and as negative PV on the upper boundary. The illumination of non-modal wave behaviour provided by Bretherton’s theory led Bishop (1993b) to extend the quantitative application of CRW theory to the non-autonomous problem of baroclinic waves undergoing horizontal deformation. Bishop found that, while normal-mode theory was unable to explain the evolution of these waves, CRW theory concisely explained a wide range of deforming baroclinic-wave evolution scenarios. It also suggested a phase diagram to summarize these scenarios graphically (Bishop 1993a). Similarly, Bishop and Thorpe (1994a,b), in considering frontal waves during frontogenesis, and Barcilon and Bishop (1998), in considering baroclinic waves in horizontal shear, found understanding of these inherently non-modal wave regimes via quantitative extensions of Bretherton’s theory.

CRW theory is most simply explained, however, in the framework of autonomous systems that do support normal-mode solutions. In this context, Davies and Bishop (1994) extended the quantitative CRW description to the Eady problem, and Heifetz *et al.* (2000) applied it to the barotropic model of Rayleigh (1880). There, the CRWs are edge waves on the two boundaries and there are no PV gradients in the interior. Robinson (1989) used quasi-geostrophic PV fluxes to show how the normal modes of the Charney (1947) model depend on interaction between surface thermal anomalies and a thin layer of PV anomalies centred just above the steering level, suggestive of a CRW extension to this problem.

The main question addressed here is whether the CRW perspective of baroclinic instability is just an attractive qualitative picture that can be made quantitative in very special cases, or whether it has more general quantitative validity. Baines and Mitsudera (1994) showed that the phase-locking mechanism between two CRW edge waves is responsible for instability in a general class of shear profiles, provided that the velocity profile is symmetric, the two CRWs being located at the edges of the unstable region.

Here, we consider perturbations to a general flow $U(y, z)$ governed by the horizontal advection of a PV that can be inverted to obtain the meridional flow. It is shown that there exists a class of perturbations that can be described by the interaction of two CRWs. This class of perturbation is the one that can be generated by the growing and decaying normal-mode subset, which is also referred to as the discrete spectrum.

Two methods are presented for determining the CRWs from the (discrete spectrum) normal modes. The ‘home-base method’ (section 2) draws a direct analogy with the ‘PV-thinking’ picture given by Hoskins *et al.* (1985), and makes those arguments quantitative. However, the home-base method involves a subjective choice; it is unclear where the home-base location should be in flows such as barotropic jets in which like-signed PV gradients are spatially separated. This lack of objectivity is not suffered by ‘the orthogonality method’ presented in section 3. There, orthogonality constraints with respect to well known globally conserved properties for disturbances to steady zonally-symmetric states are used to define the CRWs. CRW evolution equations can be derived via global integrals of the displacement form of the PV equation (subsection 3(c)), or from the Hamiltonian equations for the conservative system where the CRWs are the components (appendix). It is shown that the orthogonality method yields the same form for the CRW evolution equations as the home-base method but with different expressions for the coefficients describing interaction. These coefficients have identical values in circumstances when home-bases can be appropriately defined for the CRWs determined from the orthogonality method. Readers who are not concerned with the details of obtaining CRWs objectively may want to skip section 3. Baroclinic growth from any initial condition (which could alternatively be described by a growing normal mode and its complex conjugate) is discussed by examining the evolution equations for a CRW pair in section 4. Some concluding comments are made in section 5.

The CRW analysis has been applied to the Charney problem by Heifetz *et al.* (2004, hereafter referred to as Part II). New insights are obtained into the unstable Charney and Green modes and into the behaviour at neutral wavelengths. Methven *et al.* (personal communication) have shown how the CRW analysis can be applied to an examination of baroclinically unstable perturbations to realistic zonal jets on the sphere, as described by the primitive equations.

2. THE HOME-BASE METHOD

(a) General CRW definition

We consider a basic flow $U(y, z)$ and assume that the perturbation problem is described by the linearized horizontal advection of PV:

$$\left(\frac{\partial}{\partial t} + U \frac{\partial}{\partial x} \right) q' + v' \frac{\partial \bar{q}}{\partial y} = 0, \quad (1)$$

where q' is the perturbation PV, \bar{q}_y is the meridional gradient of PV in the basic state and v' is the perturbation meridional wind. The equation could describe the advection of Ertel–Rossby PV on isentropic surfaces, or quasi-geostrophic PV on horizontal surfaces. It is assumed that horizontal boundaries are included using the Bretherton (1966a) δ -function technique, so that (1) applies everywhere in the domain. It is also assumed that the distribution of q' in the domain may be inverted to give v' everywhere. Using the Fourier series:

$$q' = \sum q_k e^{ikx}, \quad v' = -i \sum v_k e^{ikx}, \quad (2a,b)$$

(1) becomes

$$\left(\frac{\partial}{\partial t} + ikU \right) q_k - iv_k \frac{\partial \bar{q}}{\partial y} = 0. \quad (3)$$

If a structure is to be purely propagating, then v' must be out of phase with q' everywhere. This is possible if q' and, by inversion, v' are untilted in y and z . Since v' is $\pm\pi/2$ out of phase with q' everywhere, if q_k is chosen to be real then v_k must also be real. The minus sign in (2b) is chosen so that v_k and q_k both have the same sign if the maximum in v' is to the east of the maximum in q' , as expected for a simple cyclonic PV structure. From here on, the suffix k will be dropped. For a wave-like solution of the form $e^{ik(x-ct)}$, (3) implies that the local Rossby wave speed,

$$c = U - \frac{\gamma}{k}, \quad (4)$$

where $\gamma = (v/q)\partial\bar{q}/\partial y$. Since c is a function of y and z , the untilted structure will, in general, not be retained in time if it occurs in isolation. Nevertheless, we shall refer to it as a Rossby wave.

We seek a solution of the full problem (3) in terms of two Rossby waves:

$$q = \alpha_1(t)q_1 + \alpha_2(t)q_2, \quad v = \alpha_1(t)v_1 + \alpha_2(t)v_2, \quad (5a,b)$$

where q_1, q_2, v_1 and v_2 can all be taken to be real functions of y and z since the waves are untilted, but α_1 and α_2 are, in general, complex. We set,

$$\alpha_1 = a_1 e^{i\epsilon_1}, \quad \alpha_2 = a_2 e^{i\epsilon_2}, \quad (6a,b)$$

where $a_{1,2}$ are the amplitudes and $\epsilon_{1,2}$ are the phases of waves 1 and 2.

(b) *CRW construction from a growing normal mode using home-bases*

Now consider two such Rossby waves, each describing the PV completely at two different zonal lines, corresponding to points $P_1(y_1, z_1)$ and $P_2(y_2, z_2)$, in the y - z plane. The first Rossby wave (q_1, v_1) has non-zero PV at its home-base P_1 and zero PV at P_2 , and the second Rossby wave (q_2, v_2) has non-zero PV at its home-base P_2 and zero PV at P_1 .

A simple method for constructing the required two wave structures is from the growing and decaying normal modes, which are themselves solutions of (3). If the growing mode is

$$q_g = Q(y, z) \exp[i\{\phi(y, z) - kc_r t\} + kc_i t], \quad (7a)$$

then the decaying mode is

$$q_d = Q(y, z) \exp[i\{-\phi(y, z) - kc_r t\} - kc_i t], \quad (7b)$$

where the suffices r and i denote the real and imaginary parts. To make a structure that has zero PV signature at P_2 , where the phase structure function $\phi(y_2, z_2) = \phi_2 = \text{constant}$, consider

$$q_1 = e^{-i\phi_2} q_{g0} - e^{i\phi_2} q_{d0}, \quad (8)$$

where the suffix zero refers to the modal structures (7a) at time $t = 0$ and g and d indicate the growing and decaying components. Then

$$q_1 = 2iQ(y, z) \sin[\phi(y, z) - \phi_2]. \quad (9)$$

Hence, q_1 is an untilted structure which is zero at P_2 and will, in general, be non-zero at P_1 . Similarly, q_2 can be constructed using the difference of the growing and decaying mode structures, phase shifted by minus and plus ϕ_1 , respectively. According to (4) the Rossby wave speed for wave i at its home-base is:

$$c_i^i = U^i - \frac{\gamma_i^i}{k}, \quad (10)$$

where $\gamma_i^i/k = \{(v_i/q_i)\partial\bar{q}/\partial y\}^i/k$ is the propagation speed of wave i relative to the zonal-mean flow at its home-base. Here, and subsequently, the superfix refers to the level at which the quantity is evaluated and the suffix refers to the wave. In general c_1^1 and c_2^2 will be different.

Any initial conditions that can be written in terms of q_1 and q_2 can then be written as a sum of q_g and q_d . These develop in time according to their normal-mode propagation and growth or decay. At any subsequent time, the q_g and q_d components can each be written in terms of q_1 and q_2 . Hence, the solution can be written as a linear sum of q_1 and q_2 in the form (5a,b) for all time. Substituting (5a,b) and (6a,b) into the basic equation (3) at P_1 and P_2 gives the CRW evolution equations:

$$\dot{a}_1 = -\gamma_2^1 a_2 \sin \epsilon, \quad \dot{a}_2 = \gamma_1^2 a_1 \sin \epsilon, \quad (11a,b)$$

$$\dot{\epsilon}_1 = -kc_1^1 + \gamma_2^1 \frac{a_2}{a_1} \cos \epsilon, \quad \dot{\epsilon}_2 = -kc_2^2 + \gamma_1^2 \frac{a_1}{a_2} \cos \epsilon. \quad (12a,b)$$

Here, $\epsilon = \epsilon_2 - \epsilon_1$ is the phase difference and

$$\gamma_j^i = \left(\frac{v_j}{q_i} \frac{\partial \bar{q}}{\partial y} \right)^i, \quad (13)$$

for $i, j = 1, 2$. The interaction coefficient γ_2^1 is associated with the effect on the PV at P_1 due to the meridional advection induced by wave 2, even though the latter has zero PV there. Similarly, γ_1^2 represents the effect on the PV at P_2 due to wave 1. Equations (10)–(13) are a generalization of the equations derived for the Eady problem by Davies and Bishop (1994). As pointed out by them, these equations give the time invariants,

$$\frac{1}{2\gamma_2^1} a_1^2 + \frac{1}{2\gamma_1^2} a_2^2 = \text{const}, \quad (14a)$$

$$\frac{1}{2\gamma_2^1} a_1^2 - \frac{1}{2\gamma_1^2} a_2^2 - \frac{2}{\mu} a_1 a_2 \cos \epsilon = \text{const}. \quad (14b)$$

At this stage it has only been proved that, if a solution of the form (5a,b) is to exist, then the results (10)–(14b) follow from the satisfaction of the basic equation (3) at P_1 and P_2 . However, it has not been shown that (3) is satisfied everywhere. Substituting (5a,b) and (6a,b) into (3) and using the definitions (4) and (10), the form of solution (5a,b) is consistent if everywhere,

$$(c_1 - c_1^1)q_1 = -\frac{\gamma_1^2}{k}q_2, \quad (c_2 - c_2^2)q_2 = -\frac{\gamma_2^1}{k}q_1. \quad (15a,b)$$

These equations express the conditions that the differential advection of each wave is balanced by the interaction with the other so that it retains its untilted structure. It can be confirmed analytically that the q_1 and q_2 constructed from the normal modes in this

manner do indeed satisfy the coupled pair of equations (15a,b). Further, apart from amplitude scaling, the construction is unique for initial conditions that can be expressed in terms of the growing and decaying normal-mode pair and for given P_1 and P_2 .

The pair of coupled equations (15a,b) contains four unknown constants: the values of γ_j^i , for $i, j = 1, 2$. If the PV perturbations are caused solely by meridional advection of PV, then the meridional displacements can be defined as

$$\eta_i = -q_i \left(\frac{\partial y}{\partial \bar{q}} \right), \quad (16)$$

and the four constants can be written

$$\gamma_j^i = \left(-\frac{v_j}{\eta_i} \right)^i. \quad (17a)$$

Alternatively, in terms of PV fluxes,

$$\gamma_j^i = \frac{\left(\rho v_j q_i \frac{\partial \bar{q}}{\partial y} \right)^i}{(\rho q_i^2)^i}. \quad (17b)$$

Both forms for γ_j^i show how it expresses the normalized influence of wave j on wave i at P_i . The form (17b) stresses the link between the present discussion and that by Robinson (1989), who used such PV fluxes to show how Charney modes can be viewed as mutually reinforcing interactions between the disturbance PV at different levels.

The strength of the home-base method is its direct encapsulation of the PV-thinking perspective given in the introduction and in Fig. 1. However, unless the PV gradients are concentrated in two distinct regions, there is an arbitrariness about the choice of the two home-bases. For a flow independent of y and with significant baroclinicity at the surface, it is natural to take P_1 to be at this level. This allows the construction of wave 2. As is discussed in Part II, there is not great sensitivity to the exact level chosen for P_2 when analysing Charney waves. However, in principle the home-base method does have a weakness in the subjective choice of the home-bases and the related indeterminacy in phase-locked angle ϵ and, therefore, whether the CRWs are mutually helping or hindering their counter-propagation. Hence, in the next section we present a more robust method to determine the CRWs which eliminates the subjective choice of home-bases.

3. THE ORTHOGONALITY METHOD

(a) Pseudomomentum orthogonality

In this section the analysis will be restricted to the quasi-geostrophic problem. The generalization to disturbances described by primitive-equation dynamics has been made by Methven *et al.* (personal communication). As shown by Bretherton (1966a), when all PV anomalies arise from displacing the PV gradient or boundary potential-temperature gradient, the volume-averaged meridional eddy PV flux vanishes,

$$\iint \overline{v'q'\bar{\rho}} \, dy \, dz = 0, \quad (18)$$

where the overbar represents a zonal average. It is useful to define an inner product for wave-like quantities $p' = \text{Re}\{p_k e^{ikx}\}$:

$$\langle p', q' \rangle \equiv \iint \overline{p'q'\bar{\rho}} \, dy \, dz = \iint \frac{1}{4} (p_k^* q_k + p_k q_k^*) \bar{\rho} \, dy \, dz, \quad (19)$$

where the asterisk indicates complex conjugate. For a purely propagating untilted CRW, v and q are $\pi/2$ out of phase, and therefore the zonally averaged contribution to the meridional PV flux vanishes. Thus, using the definitions (2a,b), (5a,b) and (6a,b), (18) becomes

$$a_1 a_2 \sin \epsilon \{ \langle v_2, q_1 \rangle - \langle v_1, q_2 \rangle \} = 0. \quad (20)$$

Hence, the contributions to the integrated meridional PV fluxes that each untilted CRW induces on the other are equal. Using the meridional displacement definition (16), then (18) can be written as

$$-\frac{1}{2} \frac{\partial}{\partial t} \left\langle \eta, \frac{\partial \bar{q}}{\partial y} \eta \right\rangle \equiv -\frac{\partial \mathcal{P}}{\partial t} = 0, \quad (21)$$

where \mathcal{P} is the wave activity or the pseudomomentum. The sign used here for \mathcal{P} is that of Held (1985), and the negative of that used in the definition by Andrews and McIntyre (1976). Note that extreme values of \mathcal{P} correspond to regions where both wave amplitude and PV gradient are large. Our illustration of CRWs (Fig. 1) suggests that the upper and lower CRWs have large positive and negative \mathcal{P} respectively. As pointed out by Bretherton (1966a), the conservation of \mathcal{P} yields immediately the Charney and Stern (1962) condition for instability; \bar{q}_y must change sign within the domain in order for wave growth.

Partitioning the meridional displacement into the two CRWs:

$$\eta = a_1(t) e^{i\epsilon_1(t)} \eta_1 + a_2(t) e^{i\epsilon_2(t)} \eta_2, \quad (22)$$

and defining the pseudomomentum bracket

$$\mathcal{P}_{ij} \equiv \left\langle \eta_i, \frac{\partial \bar{q}}{\partial y} \eta_j \right\rangle, \quad (23)$$

then $\mathcal{P}_{ij} = \mathcal{P}_{ji}$, since η_1 and η_2 are real by definition. Note that the Eady edge waves defined by Davies and Bishop (1994) are orthogonal with respect to the pseudomomentum norm, because each edge wave has zero meridional displacement on its opposing boundary and because \bar{q}_y is non-zero only at the two boundaries. If we impose pseudomomentum orthogonality for the general CRWs, i.e.

$$\mathcal{P}_{12} = \mathcal{P}_{21} = 0, \quad (24)$$

then the invariant total pseudomomentum can be written as

$$\mathcal{P} = \frac{\mathcal{P}_{11}}{2} a_1^2 + \frac{\mathcal{P}_{22}}{2} a_2^2 = \text{const}, \quad (25)$$

which is analogous to (14a). In the Eady model, the mean PV gradient has opposite signs on the two boundaries, and therefore \mathcal{P}_{11} and \mathcal{P}_{22} have opposite signs. (25) indicates that, in the general case, synchronous growth of the two CRWs is possible only if the spatial correlations between the magnitude of each CRW displacement and the mean PV gradient have opposite signs. Thus, (25) can be regarded as the CRW version of the Charney and Stern condition.

Since two independent orthogonality relations are required in order to determine both CRW phases, we next seek an orthogonality relation which makes the Fjørtoft (1951) condition for instability transparent in terms of CRW interaction.

(b) *Energy-growth orthogonality*

The global eddy-energy growth can be written (e.g. Pedlosky 1987) as

$$\frac{\partial E}{\partial t} = \langle iv, Uq \rangle = \langle iv, (U - U_m)q \rangle, \quad (26)$$

where the global eddy energy is

$$E = \frac{1}{2} \left(\langle u, u \rangle + \langle v, v \rangle + \left\langle \frac{f}{N}\theta, \frac{f}{N}\theta \right\rangle \right), \quad (27)$$

f is the Coriolis parameter, N is the Brunt–Väisälä frequency, θ is the scaled perturbation potential temperature and U_m is some constant reference wind. For the CRW partition, (26) becomes

$$\frac{\partial E}{\partial t} = a_1 a_2 \sin \epsilon \{ \langle v_1, (U - U_m)q_2 \rangle - \langle v_2, (U - U_m)q_1 \rangle \}. \quad (28)$$

Using (20) in the right-hand side of (28) verifies that the eddy-energy growth is invariant to the Galilean transformation by U_m . (28) indicates that, for positive vertical shear, if on average the PV structure of wave 2 is concentrated above the PV structure of wave 1, then a configuration that is favourable for energy growth is when wave 2 located to the west of wave 1 ($\sin \epsilon > 0$).

Considering the perturbation PV to arise solely from advection of the mean PV, as described in (16), then (26) can be written as,

$$\frac{\partial E}{\partial t} = \frac{1}{2} \frac{\partial}{\partial t} \left\langle \eta, (U - U_m) \frac{\partial \bar{q}}{\partial y} \eta \right\rangle, \quad (29)$$

which leads, using (21), to the conservation of pseudoenergy,

$$\mathcal{H} \equiv E - \frac{1}{2} \left\langle \eta, U \frac{\partial \bar{q}}{\partial y} \eta \right\rangle = \text{const.} \quad (30)$$

Defining, for the CRW partition,

$$\mathcal{S}_{ij} \equiv \left\langle \eta_i, (U - U_m) \frac{\partial \bar{q}}{\partial y} \eta_j \right\rangle = \mathcal{W}_{ij} - U_m \mathcal{P}_{ij}, \quad (31a)$$

$$\mathcal{W}_{ij} \equiv \left\langle \eta_i, U \frac{\partial \bar{q}}{\partial y} \eta_j \right\rangle, \quad (31b)$$

then it is clear that, for the Eady edge waves defined by Davies and Bishop (1994),

$$\mathcal{S}_{12} = \mathcal{S}_{21} = 0. \quad (32)$$

If we apply this orthogonality relation in the general case, then (29) becomes

$$\frac{\partial E}{\partial t} = \frac{\mathcal{S}_{11}}{2} \frac{\partial (a_1^2)}{\partial t} + \frac{\mathcal{S}_{22}}{2} \frac{\partial (a_2^2)}{\partial t}. \quad (33)$$

Hence, although each CRW enables the other to grow, condition (32) allows the total-energy growth to be composed of separate contributions from the two CRWs. However, it should be noted that (32) does not imply orthogonality with respect to the energy norm itself. (33) shows that growth of the CRW amplitudes lead to energy growth if the mean PV gradient and zonal wind are positively correlated in the regions where

the CRWs are concentrated. Therefore, (33) can be regarded as the CRW version of the Fjørtoft condition for instability.

Satisfying the two orthogonality conditions (24) and (32) is sufficient to determine the phases of the two CRWs, eliminating the two degrees of freedom involved with the home-base method. Furthermore, as we see in the following subsection, the constraints (24) and (32) enable the evolution equation for any given CRW to be obtained via an inner product between the CRW itself and the PV equation. Using these constraints, it is also possible to derive the CRW evolution equations directly from the conservation of pseudoenergy using Hamiltonian methods (see the appendix).

(c) *The global CRW equations*

Writing the PV equation (3) in terms of meridional displacement (16), then

$$\left(\frac{\partial}{\partial t} + U \frac{\partial}{\partial x} \right) \eta = -iv. \quad (34)$$

For the CRW partition (5a,b),(6a,b) this equation becomes:

$$\left(\frac{\partial}{\partial t} + ikU \right) (a_1 e^{i\epsilon_1} \eta_1 + a_2 e^{i\epsilon_2} \eta_2) = -i(a_1 e^{i\epsilon_1} v_1 + a_2 e^{i\epsilon_2} v_2). \quad (35)$$

Multiplying both sides of (35) by $\bar{\rho} \eta_1 \bar{q}_y$ and integrating over the whole volume, and using the orthogonality relations (24) and (32), we obtain:

$$\mathcal{P}_{11}(\dot{a}_1 + ia_1 \dot{\epsilon}_1) + ik \mathcal{W}_{11} \dot{a}_1 = -i \mathcal{M}_{11} a_1 - i \mathcal{M}_{12} a_2 e^{i\epsilon}, \quad (36)$$

where here we define

$$\mathcal{M}_{ij} \equiv \left\langle \eta_i, \frac{\partial \bar{q}}{\partial y} v_j \right\rangle. \quad (37)$$

Then, extracting the real and the imaginary parts of (36) yields, respectively,

$$\dot{a}_1 = \left(\frac{\mathcal{M}_{12}}{\mathcal{P}_{11}} \right) a_2 \sin \epsilon, \quad (38a)$$

$$\dot{\epsilon}_1 = -k \left\{ \left(\frac{\mathcal{W}_{11}}{\mathcal{P}_{11}} \right) + \left(\frac{\mathcal{M}_{11}}{k \mathcal{P}_{11}} \right) \right\} - \left(\frac{\mathcal{M}_{12}}{\mathcal{P}_{11}} \right) \frac{a_2}{a_1} \cos \epsilon. \quad (38b)$$

Recalling the form (17a) for γ_j^i in the home-base method, we define

$$\overline{\gamma}_j^i \equiv -\frac{\mathcal{M}_{ij}}{\mathcal{P}_{ii}} = \frac{\left\langle \eta_i, -\frac{\partial \bar{q}}{\partial y} v_j \right\rangle}{\left\langle \eta_i, \frac{\partial \bar{q}}{\partial y} \eta_i \right\rangle}, \quad (39a)$$

so that it has the same meaning as γ_j^i , but in a globally averaged sense. (A.5) shows that it is also proportional to the interaction energy. Similarly, defining

$$\overline{U}^i \equiv \frac{\mathcal{W}_{ii}}{\mathcal{P}_{ii}}, \quad (39b)$$

then the self-propagation speed of CRW 1 can be defined as

$$\overline{c}_1^1 = \overline{U}^1 - \overline{\gamma}_1^1/k. \quad (39c)$$

Substituting (39) into (38) we obtain

$$\dot{a}_1 = -\overline{\gamma_2^1} a_2 \sin \epsilon, \quad \dot{\epsilon}_1 = -kc_1^1 + \overline{\gamma_2^1} \frac{a_2}{a_1} \cos \epsilon. \quad (40a,b)$$

Multiplying both sides of (35) by $\overline{\rho} \eta_2 \overline{q}_y$, and integrating globally gives:

$$\dot{a}_2 = \overline{\gamma_1^2} a_1 \sin \epsilon, \quad \dot{\epsilon}_2 = -kc_2^2 + \overline{\gamma_1^2} \frac{a_1}{a_2} \cos \epsilon. \quad (40c,d)$$

We denote equation set (40a,b,c,d) as the global CRW evolution equations. Although they are homomorphic to the home-base CRW equations (11a,b) and (12a,b), note that their derivation did not include any subjective choice for the home-base locations. Nevertheless, since the zonal winds (39b) play the same role as the home-base winds in (11a,b) and (12a,b), they may be viewed as home-base winds although they are actually weighted integrals of winds at many locations.

Note that the integral performed to obtain (36), and hence the CRW equations (40a,b), from (35) is equivalent to evaluating the inner product of the PV conservation equation (34) with the CRW structure function η_1 . Hence, from a geometric perspective, the evolution equation for CRW 1 is obtained from the PV equation (34) by finding its component (or projection) in the direction given by η_1 . This is possible if and only if η_1 and η_2 satisfy (24) and (32).

Interestingly, there is nothing in the constraints (24) and (32) that demands that η_1 and η_2 be constructed from growing and decaying normal-mode pairs. Thus, in principle, the orthogonality method allows CRW-type equations to be constructed to describe the *initial* evolution of *any* two linearly-independent untilted structures. The key advantage in using complex-conjugate normal modes as a basis for CRWs is that any initial conditions that can be described entirely in terms of two such normal-mode CRWs will evolve into a structure that can still be described entirely by the same two CRWs. In contrast, initial conditions composed of two non-normal-mode CRWs will evolve into a structure that cannot, in general, be described entirely by the same CRWs. Thus, using a normal-mode basis simplifies matters considerably and it is all we consider in this paper.

Recall that the integrated PV fluxes that each CRW induces on the other are equal. Therefore, it may be shown that $\mathcal{M}_{12} = \mathcal{M}_{21}$ (using (20) and (37)). Additionally, if we choose to renormalize the CRWs with respect to pseudomomentum:

$$\mathcal{P}_{11} = -\mathcal{P}_{22}, \quad (41)$$

then the interaction coefficients are equal and opposite:

$$-\overline{\gamma_2^1} = \overline{\gamma_1^2} = \overline{\sigma}. \quad (42)$$

In order to complete the symmetry between the two CRWs, we can define the reference wind to be:

$$U_m = \frac{1}{2}(\overline{U^1} + \overline{U^2}) = \frac{1}{2} \left(\frac{\mathcal{W}_{11}}{\mathcal{P}_{11}} + \frac{\mathcal{W}_{22}}{\mathcal{P}_{22}} \right) \quad (43)$$

so that the magnitudes of the two CRWs with respect to the energy-growth norm (31a) become equal to each other; that is,

$$\mathcal{S}_{11} = \mathcal{S}_{22} = \frac{1}{2}(\mathcal{W}_{11} + \mathcal{W}_{22}). \quad (44)$$

Consequently, in the case of the Eady model, the linear combination of normal-mode structures that satisfy the orthogonality constraints given above are the same as the linear combination given by Davies and Bishop (1994). In addition, the orthogonality method extracts home-base winds (39b) at the boundaries of the Eady model. For a general zonal flow, it can be shown that the orthogonality method yields two CRWs that represent extreme values of negative and positive pseudomomentum given that each CRW has the same amplitude with respect to the energy growth norm (31a).

The orthogonality method provides an objective means to determine the CRW phases from the normal-mode structures. In time, any disturbance consisting of a growing normal mode and its complex conjugate will eventually be dominated by the growing normal mode. Alternatively, this behaviour can be described by a CRW pair that tend towards a phase-locked position. By expressing the left-hand side of (5a,b) in terms of the real and imaginary components of the growing normal-mode structure, a relationship between these components and the phase-locked CRWs is determined. This relation can be inverted to express each CRW structure (η_i) as a combination of the real and imaginary parts of the growing normal-mode displacement structure (i.e. $\eta_i = b_i(\eta_r + e_i\eta_i)$ for $i = 1, 2$, where b_i and e_i are real coefficients). Together, the orthogonality relations (24) and (32) are sufficient to determine coefficients e_1 and e_2 in terms of the inner products \mathcal{P}_{rr} , \mathcal{P}_{ri} , \mathcal{P}_{ii} , \mathcal{S}_{rr} , \mathcal{S}_{ri} and \mathcal{S}_{ii} . It can be readily shown that $\tan \epsilon_1 = -1/e_2$, $\tan \epsilon_2 = -1/e_1$ and the phase-locked angle for the growing normal mode is $\epsilon_+ = \epsilon_2 - \epsilon_1$.

4. THE CRW PERSPECTIVE ON BAROCLINIC GROWTH

(a) *Summary of the two methods for obtaining CRWs*

We have shown that baroclinic (or barotropic) growth on any zonally-symmetric state (in thermal-wind balance) can be considered in terms of the growth and propagation of two zonally-untitled waves, in circumstances where the initial disturbance consists of a linear combination of a growing normal mode and its decaying complex conjugate. The basic definitions of the wave structures were given in section 2(a); these structures do not change with time other than in amplitude and phase. Meridional wind is out of phase with PV in such a way that a positive PV anomaly in an isolated zonal wave-train would be associated with cyclonic circulation. Two such waves are referred to as ‘counter-propagating Rossby waves’ (or CRWs).

In sections 2 and 3, two methods have been presented to infer the CRWs’ phases when locked into the growing normal-mode structure. This is sufficient to determine the structures of both CRWs. Both methods also result in evolution equations for the CRW phases and amplitudes which have essentially the same form, but with different expressions for the interaction coefficients γ_j^i . The home-base method (section 2(b)) relies on the CRWs having maximum wave activity at two distinct locations in the latitude–height plane—the first located where the PV gradient is negative and the second where it is positive. Once these home-bases have been determined, the CRW structures can be calculated from (9) and their evolution can be discussed fully in terms of the PV equation and meridional wind induced at each of these home-bases without further consideration of the rest of the domain. The schematic Fig. 1 and associated PV-thinking arguments on mutual growth and relative propagation then apply rigorously to the CRWs defined using their home-bases.

The drawback of the home-base method is that suitable home-base locations must be anticipated from the basic state before the CRW structures are known. The orthogonality method provides an objective means to determine the CRW phases from the

growing normal-mode structure. The orthogonality conditions on CRW pseudomomentum (24) and energy growth (32) are sufficient to determine their phase-locked positions, ϵ_1 and ϵ_2 , and also their structures, although amplitudes are only determined uniquely by the renormalization (41). However, the interaction coefficients obtained by this method involve global integrals over the CRW structures which are harder to interpret than the local expressions obtained by the home-base method.

It is important to note that the evolution of a CRW pair obtained by the orthogonality method can be described by the home-base CRW equations (10)–(13), provided that the following home-bases can be defined. If a point exists where the zero PV contour of wave 1, $q_1(y, z) = 0$, intersects the zonal-wind contour $U(y, z) = \overline{U^2}$ in the meridional plane (where $\overline{U^2}$ is the mean wind (39b)), it can define the home-base of wave 2, P_2 . Similarly, if $q_2(y, z) = 0$ intersects $U(y, z) = \overline{U^1}$, this point can define P_1 . The CRW phases, structures and phase speeds obtained by the home-base method for this particular P_1 and P_2 would be identical to those from the orthogonality method.

(b) Baroclinic development

A normal-mode description of the development from initial conditions composed of a growing normal mode and its decaying complex conjugate has the advantage of simple temporal behaviour of time-independent structures. An advantage for the interacting Rossby-wave description is that the structures are typically much simpler, being untilted in both y and z . Further, the behaviour can be understood readily in PV terms by considering the propagation and interaction of waves at just two locations, the home-bases P_1 and P_2 .

In the following discussion, it will be assumed that the PV in wave 1 is characterized by its distribution at P_1 in the sense that inversion gives the expected cyclonic circulation corresponding to positive PV there, and further that the circulation has the same phase at P_2 . This was the assumption made in the introduction for the discussion of Fig. 1. A corresponding assumption is also made for the second wave. If q_1^1 and q_2^2 are taken to be positive, then using (2) these assumptions amount to stating that v_1^1, v_1^2, v_2^2 and v_2^1 are all positive.

Equation (4) then gives the usual sense of Rossby-wave propagation, that each wave's speed is westward or eastward relative to the zonal flow at its base, depending on whether \overline{q}_y is positive or negative there. From (13), the sign of the interaction coefficients γ_2^1 and γ_1^2 is that of \overline{q}_y at P_1 and P_2 , respectively. From (11a,b), the growth of both waves implies that γ_2^1 and γ_1^2 have opposite signs. This implies that \overline{q}_y must have opposite signs at P_1 and P_2 which is a local version of Charney and Stern (1962) condition for instability.

By imposing the two orthogonality constraints, or defining two home-base points, the structure of the lower wave (q_1, v_1) is determined up to some arbitrary constant factor, b_1 , and the upper wave (q_2, v_2) is determined up to a different factor, b_2 . There is freedom to renormalize the amplitudes so that their interaction coefficients are of equal magnitude, σ , but opposite in sign (42), which in turn determines the ratio b_1/b_2 . For orthogonal CRWs this is equivalent to specifying that they have equal and opposite pseudomomentum (41). Throughout, we refer to the wave with negative pseudomomentum as the 'lower wave' since it must have greatest wave activity near the ground where the PV gradient (δ function) is negative (therefore γ_2^1 and γ_1^1 are negative). The 'upper wave' has positive pseudomomentum and, therefore, exists where the positive interior PV gradient is large (thus γ_1^2 and γ_2^2 are positive). From (10), the upper wave moves westward relative to the flow at P_2 and the lower wave moves

eastward relative to the flow at P_1 . In this situation the CRW equations (11a,b) and (12a,b) (or from orthogonality (40a,b,c,d)) reduce to

$$\dot{\chi} = \sigma(1 - \chi^2) \sin \epsilon, \quad \dot{\epsilon} = \mu + \sigma \left(\chi + \frac{1}{\chi} \right) \cos \epsilon \quad (45a,b)$$

where χ is the amplitude ratio a_1/a_2 and

$$\mu = -k(c_2^2 - c_1^1) \quad (46)$$

is proportional to the difference in Rossby-wave speeds associated with self-induced propagation. (45a,b) shows that the evolution of any initial-value problem of (3) that can be described in terms of two interacting CRWs depends solely on their initial amplitude ratio χ_0 and initial phase difference ϵ_0 .

Following Davies and Bishop (1994), we first consider synchronized growth, $\chi = 1$. From (45a) it is clear that, if initially $\chi = 1$, it will remain so. Note that, because of renormalization (42), $\chi = 1$ does not imply that the two waves have the same PV amplitude at their respective home-bases. Nevertheless, we call a_1, a_2 the CRW amplitudes. In this case (11a,b) and (12a,b) reduce to

$$\dot{a} = \sigma a \sin \epsilon, \quad \dot{\epsilon} = \mu + 2\sigma \cos \epsilon. \quad (47a,b)$$

The growing and decaying normal modes are the synchronized growth solutions with $\dot{\epsilon} = 0$, and so the relative phase angle for phase locking, ϵ_+ , can be expressed as:

$$\cos \epsilon_+ = -\mu/2\sigma = \frac{k(c_2^2 - c_1^1)}{2(-\gamma_2^1 \gamma_1^2)^{1/2}}. \quad (48)$$

This equation has solutions only if the modulus of the right-hand side is less than or equal to unity. The difference in Rossby-wave speeds must be small enough and the mutual interaction large enough for phase locking to occur. From (10), this condition may be written

$$|k(U^2 - U^1) - \gamma_2^2 + \gamma_1^1| < 2(-\gamma_2^1 \gamma_1^2)^{1/2}. \quad (49)$$

Assuming that the meridional wind induced by each wave is stronger at its home-base than at the other base, or equivalently (since $\gamma_2^1 = -\gamma_1^2$) the wind induced by itself at its home-base is stronger than the wind induced there by the other CRW, we can write the following relations:

$$-\gamma_1^1 > -\gamma_2^1 > 0 \quad \text{and} \quad \gamma_2^2 > \gamma_1^2 > 0. \quad (50)$$

In this case (49) implies that

$$\{(-\gamma_1^1)^{1/2} - (\gamma_2^2)^{1/2}\}^2 < k(U^2 - U^1) < \{(-\gamma_1^1)^{1/2} + (\gamma_2^2)^{1/2}\}^2. \quad (51)$$

Hence, a necessary condition for normal-mode solutions is that $(U^2 - U^1)$ is positive. This is clearly related to the Fjørtoft (1951) condition that \bar{q}_y and U must be positively correlated over the whole domain (since CRW 2 has the positive pseudomomentum). The inequality (51) indicates that the self-propagation speed of both CRWs should be large enough, relative to the shear, in order to resist it and maintain a phase-locked state. In a normal-mode configuration the two CRWs propagate with the same phase speed,

$$c_r = -\dot{\epsilon}_1/k = -\dot{\epsilon}_2/k, \quad (52)$$

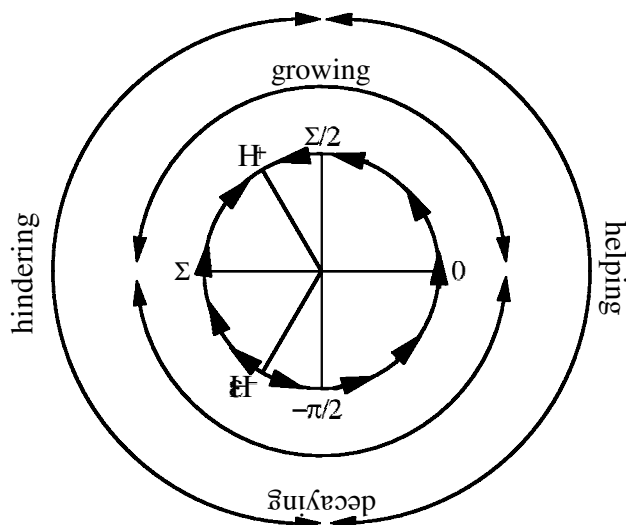


Figure 2. The relative phase angle, ϵ , in the synchronous initial-value problem. There is growth for $0 < \epsilon < \pi$ and decay for $-\pi < \epsilon < 0$. For $-\pi/2 < \epsilon < \pi/2$ each wave 'helps' the other one propagate against the relative flow at its home-base, while in the other half domain each wave 'hinders' this propagation of the other (i.e. each wave acts to move the other in the sense of the relative flow at its home-base). ϵ_+ corresponds to the growing normal-mode phase and ϵ_- to the decaying normal-mode phase. Here, the case is shown where the normal-mode phases are in the range of hindering. In the synchronous initial-value problem the phase difference changes in the sense of moving from ϵ_- to ϵ_+ , as indicated. When amplitudes are not equal, the phase behaviour can be more complicated initially.

where the subscript r denotes the real part. Thus, using (12a,b), (42) and $\chi = 1$ we obtain for the normal-mode configuration that

$$c_r = \frac{1}{2}(c_1^2 + c_2^2). \quad (53)$$

From (47a) the growth rate of the unstable normal mode is

$$\sigma_+ = \sigma \sin \epsilon_+, \quad (54)$$

where $0 < \epsilon_+ < \pi$ while, for the decaying normal mode, $\sigma_- = -\sigma_+$ and, therefore, $\epsilon_- = -\epsilon_+$. This implies that, for the growing mode, the lower wave leads the upper wave, a westward tilt with height. If the phase speed c_2^2 of the upper wave exceeds that of the lower wave c_1^2 , (48) implies that the phase difference must be between 0 and $\pi/2$. If the upper wave moves slower, then the difference must be between $\pi/2$ and π (as is the case for the CRWs of the Charney's gravest mode, where $\epsilon_+ \approx 0.79\pi$). These deductions are all consistent with Fig. 1 and the discussion there.

Figure 2 indicates the behaviour in the synchronous initial-value problem described by (47a,b). There is growth whenever $0 < \epsilon < \pi$ and decay when $-\pi < \epsilon < 0$. The maximum instantaneous growth rate is σ and occurs when the phase difference is $\pi/2$. From (54) this instantaneous growth rate will be larger than, or equal to, that of the normal mode. For $-\pi/2 < \epsilon < \pi/2$ each wave 'helps' the other to propagate against the relative flow at its home-base and, in the other half domain, each wave 'hinders' this propagation of the other, i.e. each wave acts to move the other in the sense of the relative flow at its home-base.

The phase difference always changes in the sense of moving from ϵ_- to ϵ_+ . This tendency has extrema $\dot{\epsilon} = \mu \pm 2\sigma$, when the actual phase difference is 0 and π

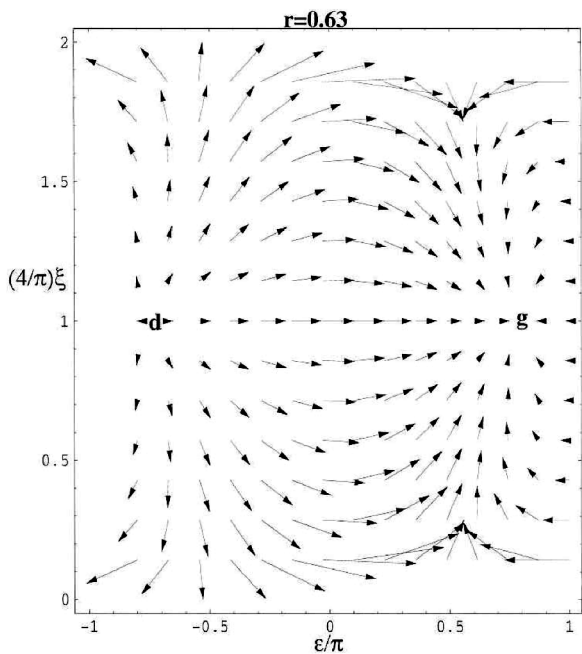


Figure 3. Phase diagram for all initial-value problems described by the interaction of the counter-propagating Rossby-wave pair obtained from superposing the fastest growing Charney mode (g) and its decaying complex conjugate (d) using the ‘home-base method’. The abscissa is the phase difference, ϵ , between the upper and the lower CRWs, scaled by π . The ordinate is the inverse tangent, ξ , of the amplitude ratio of the upper and the lower CRWs, scaled by $\pi/4$. The arrows indicate the phase-space flow where their x and y components are the time derivative of the scaled ϵ and ξ , as determined respectively by (55a,b).

and tends to zero near ϵ_- and ϵ_+ . At the maximum instantaneous growth ($\epsilon = \pi/2$), $\dot{\epsilon} = \mu$, and is associated purely with the differential Rossby-wave speeds.

As in the paper by Davies and Bishop (1994), analytic solutions can be found for both the synchronous problem and for the problem with one CRW initially zero. For the general initial-value problem solution paths can be illustrated using phase-space diagrams. In order to avoid singularity in χ for the initial condition of zero wave 2, we transform (45a,b) into

$$\dot{\xi} = \sigma \cos (2 \xi) \sin \epsilon, \quad \dot{\epsilon} = \mu + \frac{2 \sigma \cos \epsilon}{\sin 2 \xi}, \tag{55a,b}$$

where $\xi = \arctan \chi$, so that $\chi = a_1/a_2 \in (0, \infty)$ corresponds to $\xi \in (0, \pi/2)^\dagger$. Then, a compact way to visualize the evolution from any initial condition (ϵ_0, ξ_0) , can be obtained by following the flow on a phase space of (ϵ, ξ) . Figure 3 shows such a phase space for the CRWs of the most unstable Charney mode (for more details on the Charney model CRWs see Part II). Although this is a specific case, it demonstrates clearly the nature of CRW evolution. The abscissa in Fig. 3 is ϵ/π and the ordinate is ξ , scaled by $\pi/4$ so that values 0, 1, 2 of the ordinate correspond to $a_1 = 0$, $a_1 = a_2$, $a_2 = 0$, respectively. At each point (ϵ, ξ) of the phase space, an arrow is plotted with horizontal

[†] (55a,b) generalize (15a,b) of Heifetz *et al.* (2000), which are the edge-wave equations for the barotropic Rayleigh (1880) model of shear instability. In their notation γ corresponds to ξ and $\Delta \epsilon$ to ϵ . For the Rayleigh model $\sigma = (\Lambda/2) e^{-K}$ and $\mu = \Lambda(1 - K)$, where Λ and K are the normalized meridional mean shear and wave number, respectively.

and vertical components equal to the local values of ϵ and $\dot{\xi}$, respectively, which are calculated from (55a,b). Then, the evolution from any initial condition (ϵ_0, ξ_0) can be easily tracked by simply following the arrows.

For the general initial-value problem, a_1 and a_2 both still grow or decay according to the sign of $\sin \epsilon$. Hence, the right-hand side of Fig. 3 corresponds to the growing regime and the left-hand side to the decaying regime. Since, from (11a,b), the growth or decay of each CRW depends on the amplitude of the other CRW, the smaller of the waves grows or decays faster. Consequently, in the growing regime the growth tends to equalize the CRWs' amplitudes toward synchronous growth, and therefore in the right-hand side of the phase space all arrows point toward the synchronous line $\xi = \pi/4$. Similarly, decay tends to increase the amplitude ratio, and therefore on the left-hand side of Fig. 3 all arrows point away from the synchronous line. (45a) indicates that χ will tend to unity, i.e. the synchronous case, on a time-scale of order σ^{-1} . If the initial value of χ is far from unity, then (45b) indicates that the initial phase changes will be rapid and the growing and decaying normal-mode phases can be crossed. This is indicated in Fig. 3 by the arrows above and below 'g' and 'd'. For the non-synchronous case the curve of phased-locked configuration, $\dot{\epsilon} = 0$, satisfies from (55b) $\cos \epsilon = \cos \epsilon_+ / \sin 2\xi$. However (55a) indicates that then $\dot{\xi} \neq 0$. Hence, if the two CRWs approach a phase-locked state, the amplitude ratio will continue to change until it is unity at state 'g'.

The Eady model discussion by Davies and Bishop (1994) is a particular case of the general analysis presented here. The two levels z_1 and z_2 are at the two boundaries and the CRWs propagate on the boundary temperature gradients. As pointed out by Bretherton (1966b), the Eady model has a short-wave cut-off which arises when the CRW depth scale, which is proportional to wavelength, becomes too small relative to the specified separation of the boundaries. When this occurs (51) cannot be satisfied—the counter-propagation of Eady edge waves against the vertical wind shear becomes too weak to maintain phase-locking.

5. CONCLUDING REMARKS

Consider a dynamical system of two pendulums coupled by a spring. It is a simple matter to solve the eigenproblem of this system and find the two normal modes whose superposition can describe any linear motion. However, the mathematical solution does not necessarily provide a complete understanding of how the system is working. In this example, we know the independent behaviour of each pendulum under the restoring force of gravity and the elastic properties of the spring and, therefore, it is easy to combine this information and suggest a mechanistic intuitive explanation of how these three components are working together as a system.

Due to the complex nature of problems in fluid dynamics it is usually very difficult to provide such mechanistic explanations, even when eigensolutions are found mathematically. Nevertheless, Bretherton (1966b) provided such an explanation for the two-layer baroclinic instability problem. In this view, the basic components of the system are two counter-propagating Rossby waves. The propagation mechanism for each wave, as described in the introduction, is provided by a simple restoring mechanism, referred to by McIntyre (1992) as the 'Rossby elasticity'. Each wave only has PV anomalies, associated with fluid displacements in the presence of a meridional PV gradient, in one layer of the fluid. However, the circulation anomalies induced by each wave are felt in the other layer, resulting in interaction between the CRWs. The sense of interaction can be described through knowledge of the relative phase of the CRWs, the PV gradient and zonal wind in each layer. Using this approach it is straightforward to combine mentally

the information and to predict how, for example, an initially isolated upper-level disturbance could create, reinforce and lock onto a lower-level disturbance in a normal-mode configuration, resulting in their mutual growth.

The main criticism of Bretherton's explanation has been that it applies only to a two-layer model, or to a situation in which the CRWs are edge waves that must be supported on a corresponding PV gradient structure which is concentrated into delta functions on the boundaries, as in the Eady (1949) model. In the general case, however, the mean PV gradient might fill all space; normal mode solutions do exist, but it has not been clear how to generalize Bretherton's insight. The beauty and the simplicity of the CRW explanation has provided motivation for the search for a generalization that would be applicable to any given baroclinic and/or barotropic basic state which supports normal-mode instability.

We have suggested two methods of generalization. In the first, the home-base method, we followed the physical rationale behind the CRW structures in a two-layer model, by characterizing two untilted CRWs, which are constructed from a pair of growing and decaying modes, by their behaviour at two zonal lines, referred to as their home-bases. Although both CRWs fill all space we forced the PV signature of each to vanish on the home-base of the other. We could then write the CRW evolution equations on those two home-bases in an exact analogous form to the Eady edge-wave equations derived by Davies and Bishop (1994). The modal phase-locking configuration, as well as the non-modal evolution, of the discrete spectrum initial value solution, can all be interpreted in terms of simple CRW interaction which depends on the CRWs' amplitude ratio, phase difference, interaction coefficient and the difference in wave speeds. Furthermore, local versions of the global integral conditions for instability of Charney and Stern (1962) and Fjørtoft (1951) are obtained, suggesting a mechanistic CRW explanation for why a change in the sign of the mean PV gradient within the domain, and a positive correlation between the mean PV gradient and the mean zonal wind, are necessary to obtain instability.

Although we demonstrated that a CRW description can indeed be applied for a general baroclinic basic state, the home-base method suffers from the need to define two home-base locations, the choice of which is in general subjective. If the mean PV gradient is mainly concentrated into two layers with opposite signs (for example, a negative PV gradient at the surface associated with temperature gradients and a positive PV gradient at the tropopause), then this method can be applied straightforwardly by locating the home-bases in the centres of the two layers. However, for a more realistic PV gradient, such as the case of baroclinic jet, it is not clear where to locate the home-bases.

Thus, in order to make the CRW description more robust, we suggested an alternative method—the 'orthogonality method'—which eliminates the two subjective choices by making the CRWs orthogonal with respect to wave activity (pseudomomentum) and orthonormal with respect to energy growth. The motivation for this method came from noting that the Eady edge waves satisfy such conditions. The pseudomomentum orthogonality constraint (24) is the only one enabling CRW evolution equations to be derived, either from the global inner product of the displacement form of the PV equation with the PV structure of each CRW (see section 3(c)) or by Hamiltonian methods where the CRWs are the system components and the Hamiltonian is the pseudoenergy (see the appendix). Enforcing orthogonality of the shear term (32) as the second constraint is the only means to obtain an interaction coefficient, σ , which depends solely upon the PV and meridional wind structures of the CRWs and PV gradient, as was found with the home-base method. The disadvantage of the orthogonality method is that it does not

relate so directly to the PV-thinking analysis of the introduction and Fig. 1, because the interaction coefficients are evaluated by global integrals rather than local properties at home-bases. However this can sometimes be recovered when home-bases can be determined from the orthogonal CRWs.

It should be emphasized that the CRW structures discussed here, in both methods, are constructed from normal modes. Therefore, in contrast to the Eady model where the edge waves can be constructed independently, here in the general case one must solve a priori the eigenproblem and find the normal modes in order to obtain the CRWs. Thus, the CRWs should be regarded as an analysis tool which provides some fundamental insight into the baroclinic growth mechanism of the discrete spectrum solution.

CRWs cannot address those aspects of the initial-value problem related to the continuous spectrum (Pedlosky 1964; Farrell 1982). However, Badger and Hoskins (2001) have shown how the behaviour of rapidly growing initial structures can be analysed in PV terms, and that they quickly develop into near normal-mode-like structures. Thus, CRWs together with Badger and Hoskins (2001) work may be considered as giving a basic PV framework for the discussion of the linear growth of mid-latitude weather systems.

In Part II we apply the two CRW methods to the Charney (1947) model. We choose the Charney model as an example since its basic state is quite similar to the Eady model, yet it contains only one boundary and a mean PV gradient which fills all space. It is found that, for Charney modes, the CRWs derived by the two methods are in fact very similar. Therefore, the PV-thinking explanation for baroclinic growth, afforded by the association of Fig. 1 with the CRWs obtained by the home-base method, applies to a good approximation to CRWs obtained using the orthogonality method. This highlights the parallels between baroclinic growth on the Eady and Charney basic states. It is shown that the CRW perspective provides new insights into the behaviour of perturbations in the Charney problem. Methven *al.* (personal communication) have obtained the CRWs from normal-mode disturbances to realistic zonal jets described by the primitive equations on the sphere. Some aspects of the nonlinear evolution of these baroclinic waves are also explained by the CRW analysis.

ACKNOWLEDGEMENTS

Craig Bishop acknowledges support from National Science Foundation grants ATM-96-12502 and ATM-98-14376. Support from the Office of Naval Research grant N00014-00-1-0106 and Program Element 0601153N Project number BE-0333-0345 are also gratefully acknowledged. John Methven is grateful for an Advanced Fellowship sponsored jointly by the Natural Environment Research Council and the Environment Agency. All authors thank the two anonymous referees for their helpful reviews.

APPENDIX

Hamiltonian derivation of the CRW evolution equations

If a CRW pair is orthogonal with respect to pseudomomentum then the total pseudomomentum is composed of separate contributions from each CRW:

$$p_1 = a_1^2 \mathcal{P}_{11}/2; \quad p_2 = a_2^2 \mathcal{P}_{22}/2 \quad (\text{A.1})$$

as seen from (25). The CRWs can also be assigned positions:

$$x_1 = \epsilon_1/k; \quad x_2 = \epsilon_2/k. \quad (\text{A.2})$$

The total pseudoenergy (30) for the pair of CRWs can be written:

$$\mathcal{H} = a_1^2(E_{11} - \mathcal{W}_{11}/2) + a_2^2(E_{22} - \mathcal{W}_{22}/2) + 2a_1a_2(E_{12} - \mathcal{W}_{12}/2) \cos \epsilon \quad (\text{A.3})$$

where E_{ij} denotes disturbance energies and \mathcal{W}_{ij} is the shear term (31b). Now consider the conservative system where the Hamiltonian is the pseudoenergy (A.3) and the CRWs are the components. The Hamiltonian equations:

$$\frac{\partial p_i}{\partial t} = -\frac{\partial \mathcal{H}}{\partial x_i}; \quad \frac{\partial x_i}{\partial t} = \frac{\partial \mathcal{H}}{\partial p_i} \quad (\text{A.4})$$

reduce to the CRW evolution equations (40a,b,c,d), where the interaction coefficients are given by $\gamma_2^1 = k(2E_{12} - \mathcal{W}_{12})/\mathcal{P}_{11}$ and $\gamma_1^2 = k(2E_{12} - \mathcal{W}_{12})/\mathcal{P}_{22}$. The CRW self-induced phase speeds are $\overline{c_1^1} = (\mathcal{W}_{11} - 2E_{11})/\mathcal{P}_{11}$ and $\overline{c_2^2} = (\mathcal{W}_{22} - 2E_{22})/\mathcal{P}_{22}$. For quasi-geostrophic dynamics it can be shown that the disturbance energy $E = -\langle \psi, q \rangle / 2$, where ψ is disturbance stream function. Hence

$$E_{ij} = -\mathcal{M}_{ij}/(2k), \quad (\text{A.5})$$

showing that the interaction coefficients and self-propagation speeds obtained from the Hamiltonian are identical to (39) obtained by global integrals of the PV equation (section 3(c)), provided that the second orthogonality constraint $\mathcal{W}_{12} = \mathcal{S}_{12} = 0$ is applied. Thus orthogonal CRWs can be regarded as components of a Hamiltonian system describing disturbance evolution and their strength of interaction is proportional to the interaction energy E_{12} .

REFERENCES

- | | | |
|------------------------------------|-------|---|
| Andrews, D. G. and McIntyre, M. E. | 1976 | Planetary waves in horizontal and vertical shear: the generalized Eliassen–Palm relation and the mean zonal acceleration. <i>J. Atmos. Sci.</i> , 33 , 2031–2048 |
| Badger, J. and Hoskins, B. J. | 2001 | Simple initial value problems and mechanisms for baroclinic growth. <i>J. Atmos. Sci.</i> , 58 , 38–49 |
| Baines, P. G. and Mitsudera, H. | 1994 | On the mechanism of shear-flow instabilities. <i>J. Fluid Mech.</i> , 276 , 327–342 |
| Barcilon, A. and Bishop, C. H. | 1998 | Non-modal development of baroclinic waves undergoing horizontal shear deformation. <i>J. Atmos. Sci.</i> , 55 , 3583–3597 |
| Bishop, C. H. | 1993a | On the behaviour of baroclinic waves undergoing horizontal deformation. I: The ‘RT’ phase diagram. <i>Q. J. R. Meteorol. Soc.</i> , 119 , 221–240 |
| | 1993b | On the behaviour of baroclinic waves undergoing horizontal deformation. II: Error bound amplification, Rossby wave diagnostics. <i>Q. J. R. Meteorol. Soc.</i> , 119 , 241–268 |
| Bishop, C. H. and Thorpe, A. J. | 1994a | Frontal wave stability during moist deformation frontogenesis: Part 1. Linear wave dynamics. <i>J. Atmos. Sci.</i> , 51 , 852–873 |
| | 1994b | Frontal wave stability during moist deformation frontogenesis: Part 2. The suppression of non-linear wave development. <i>J. Atmos. Sci.</i> , 51 , 874–888 |
| Bretherton, F. P. | 1966a | Critical-layer instability in baroclinic flows. <i>Q. J. R. Meteorol. Soc.</i> , 92 , 325–334 |
| | 1966b | Baroclinic instability and the short-wave cut-off in terms of potential vorticity. <i>Q. J. R. Meteorol. Soc.</i> , 92 , 335–345 |
| Charney, J. G. | 1947 | The dynamics of long waves in a baroclinic westerly current. <i>J. Meteorol.</i> , 4 , 135–163 |
| Charney, J. G. and Stern, M. E. | 1962 | On the stability of internal baroclinic jet in a rotating atmosphere. <i>J. Atmos. Sci.</i> , 19 , 159–172 |
| Davies, H. C. and Bishop, C. H. | 1994 | Eady edge waves and rapid development. <i>J. Atmos. Sci.</i> , 51 , 1930–1946 |
| Eady, E. T. | 1949 | Long waves and cyclone waves. <i>Tellus</i> , 1 , 33–52 |

- Farrell, B. F. 1982 The initial growth of disturbances in a baroclinic flow. *J. Atmos. Sci.*, **39**, 1663–1686
- Fjørtoft, R. 1951 ‘Stability properties of large-scale atmospheric disturbances’. Pp. 454–463 in *Compendium of Meteorology*, American Meteorological Society, Boston, USA
- Heifetz, E., Bishop, C. H. and Alpert, P. 2000 Counter-propagating Rossby waves in the barotropic Rayleigh model of shear instability. *Q. J. R. Meteorol. Soc.*, **125**, 2835–2853
- Heifetz, E., Methven, J., Hoskins, B. J. and Bishop, C. H. 2004 The counter-propagating Rossby-wave perspective on baroclinic instability. II: Application to the Charney model, *Q. J. R. Meteorol. Soc.*, **130**, 233–258
- Held, I. M. 1985 Pseudomomentum and the orthogonality of modes in shear flow. *J. Atmos. Sci.*, **42**, 2280–2288
- Hoskins, B. J., McIntyre, M. E. and Robertson, A. W. 1985 On the use and significance of isentropic potential vorticity maps. *Q. J. R. Meteorol. Soc.*, **111**, 877–946
- McIntyre, M. E. 1992 ‘Atmospheric dynamics: Some fundamentals with observational implications’. Pp. 313–386 in *Proceedings of the International School of Physics ‘Enrico Fermi’ 115th course*, Eds. J. C. Gille and G. Visconti, North-Holland
- Pedlosky, J. 1964 An initial value problem in the theory of baroclinic instability. *Tellus*, **16**, 12–17
- Rayleigh, Lord 1887 *Geophysical Fluid Dynamics*, Second Edition, Springer-Verlag
- Rayleigh, Lord 1880 On the stability, or instability, of certain fluid motions. *Proc. London Math. Soc.*, **9**, 57–70
- Robinson, W. A. 1989 On the structure of potential vorticity in baroclinic instability. *Tellus*, **41A**, 275–284